International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

Advanced Telemetry Correlation Techniques for Real-Time Reliability Engineering in Edge-Cloud Systems

Pruthvi Raj Seknametla

Independent Researcher

pruthviraj.seknametla@ieee.org

Vol. 8 No. 8 (2026): IJSTC

Abstract: The rapid expansion of edge-cloud computing architecture has introduced unprecedented observability challenges arising from the geographic dispersion of workloads, heterogeneous infrastructure substrates, and bandwidth-constrained interconnection fabrics. Traditional telemetry pipelines, designed for centralized cloud environments with homogeneous connectivity, fail to deliver the cross-layer diagnostic visibility required for effective real-time reliability engineering across distributed edge and cloud tiers. This paper presents a context-aware telemetry correlation engine (CATCE) that employs semantic enrichment, adaptive signal fusion, and topology-guided causal reasoning to correlate metrics, distributed traces, structured logs, and infrastructure events across edge-cloud boundaries in real time. The framework introduces a novel context propagation protocol that preserves causal relationships even under intermittent edge connectivity, and a resource-proportional analysis strategy that distributes computational workloads according to available capacity at each tier. Evaluation on a production-representative testbed with 18 edge locations and a multi-region cloud backbone demonstrates a 71.2 percent reduction in mean fault isolation time relative to conventional centralized approaches, with sustained diagnostic accuracy of 94.6 percent under degraded network conditions. The findings offer actionable architectural guidance for reliability engineering teams operating at the edge-cloud frontier.

Keywords: Telemetry Correlation, Edge-Cloud Architecture, Real-Time Reliability Engineering, Observability, Causal

Inference, AIOps, Site Reliability Engineering

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

I. INTRODUCTION

Edge-cloud computing represents a paradigmatic shift in distributed systems architecture, extending the computational perimeter from centralized data center regions to a geographically dispersed continuum of processing nodes that range from hyperscale cloud facilities through regional aggregation points to resource-constrained far-edge locations situated at the physical boundary between digital infrastructure and the physical world. This architectural evolution has been driven by application domains whose latency, bandwidth, and data sovereignty requirements cannot be satisfied by cloud-only deployments including autonomous mobility systems, industrial process automation, immersive media streaming, and real-time financial risk computation. As enterprises transition from pilot-scale edge experiments to production-grade distributed platforms serving critical business functions, the demand for robust, real-time reliability engineering capabilities across the full edge-cloud topology has become a first-order operational concern.

Reliability engineering in cloud-native environments depends fundamentally on observability: the ability to infer the internal state of a system from the external signals it emits. The observability discipline has matured considerably in centralized cloud contexts, where metrics, distributed traces, and structured logs flow through well-provisioned collection pipelines to unified analysis platforms that can correlate signals with access to the complete system state. Edge-cloud architectures disrupt this model by introducing bandwidth constraints on telemetry collection, network partitions that interrupt signal continuity, clock synchronization challenges that complicate temporal correlation, and infrastructure heterogeneity that fragments semantic consistency across telemetry sources.

The correlation of heterogeneous telemetry signals the process of establishing meaningful relationships among observations drawn from different signal types, system components, and infrastructure tiers—constitutes the analytical foundation upon which all higher-level diagnostic reasoning depends. Without effective correlation, anomaly detection operates on isolated signal streams and cannot distinguish localized transient fluctuations from symptoms of systemic failures propagating across tier boundaries. Root-cause analysis becomes intractable when the causal chain traverse's edge-to-cloud boundaries that fragment the observability pipeline. Incident prioritization loses accuracy when the customer impact of edge-tier anomalies cannot be connected to backend service health through correlated evidence.

Existing approaches to telemetry correlation in distributed systems reveal a significant architectural mismatch with edge-cloud requirements. Centralized correlation engines assume that all telemetry is accessible at a single processing location with negligible collection latency an assumption that fails when edge-to-cloud links operate under severe bandwidth budgets and variable latency profiles. Purely decentralized approaches preserve local processing autonomy but sacrifice the cross-tier visibility essential for diagnosing failures whose symptoms and root causes reside in different architectural tiers. Neither paradigm adequately addresses the fundamental challenge of

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

maintaining diagnostic coherence across a system whose components span heterogeneous networks, disparate hardware capabilities, and inconsistent telemetry schemas.

This paper presents the Context-Aware Telemetry Correlation Engine (CATCE), a framework purpose-built for real-time reliability engineering in edge-cloud systems. CATCE introduces three principal innovations. First, a semantic enrichment layer that normalizes and contextualizes telemetry from heterogeneous edge and cloud sources into a unified analytical representation. Second, an adaptive signal fusion mechanism that dynamically adjusts the depth and breadth of cross-signal correlation based on the computational resources available at each processing tier and the diagnostic significance of current system conditions. Third, a topology-guided causal reasoning module that leverages the known structural relationships among edge sites, network interconnections, and cloud services to constrain the causal search space and accelerate fault isolation.

The remainder of this paper proceeds as follows. Section II surveys relevant prior work. Section III details the CATCE architecture and its constituent components. Section IV describes the experimental testbed and evaluation methodology. Section V presents quantitative results. Section VI offers discussion of practical implications and acknowledges limitations. Section VII concludes with a summary of contributions and future research directions.

II. LITERATURE REVIEW

A. Foundations of Distributed System Observability

The conceptual framework for modern distributed system observability traces its lineage to the seminal work on distributed tracing by Sigelman, whose Dapper system demonstrated that causally connected request paths through large service-oriented architectures could be captured, reconstructed, and analyzed to diagnose latency anomalies and fault propagation. Subsequent systems extended this foundational concept: Zipkin, Jaeger, and the OpenTelemetry standard progressively matured the tooling ecosystem, while academic work by Sambasivan and Kaldor addressed the scalability and sampling challenges inherent in tracing production systems processing billions of requests daily.

The complementary role of metric-based monitoring was formalized through the SRE discipline's emphasis on service level indicators (SLIs), service level objectives (SLOs), and error budgets as quantitative reliability governance mechanisms. Log-based analysis, particularly when applied to structured and semantically annotated log events, provides a third observability dimension that captures discrete state transitions and error conditions not easily expressed through continuous metric series or request-scoped traces. The analytical power of correlating these three signal types has been recognized in both practitioner literature and academic research, yet the practical challenge of achieving real-time correlation at scale remains substantial.

B. Telemetry Correlation Techniques

Research on automated telemetry correlation has advanced along several methodological axes. Statistical approaches, exemplified by the work of Aggarwal, employ learned embedding spaces to align metric time series and log event streams, discovering latent correlations between signal types that facilitate unified fault diagnosis. Graph-based methods, notably the work of Chen and Wu, model service dependency structures as directed graphs annotated with telemetry features, applying graph neural networks or causal discovery algorithms to localize root causes within the topology. Temporal correlation techniques, including Granger causality analysis applied by Meng et and transfer entropy methods explored by Li, exploit the temporal ordering of anomaly propagation to distinguish causes from symptoms.

A recurrent limitation across these approaches is their implicit assumption of centralized telemetry access. The correlation algorithms presuppose that all relevant signals have been collected, temporally aligned, and made available to a single analytical engine operating with sufficient computational resources. This assumption is compatible with intra-datacenter and single-region cloud deployments but becomes untenable when applied to edge-cloud topologies where telemetry arrives with variable delays, inconsistent completeness, and heterogeneous fidelity levels imposed by bandwidth and compute constraints at edge sites.

C. Edge Computing Monitoring and Reliability

Research specifically targeting edge computing observability has emerged more recently. Shi identified monitoring under resource constraints as a primary research challenge for edge computing, noting that conventional monitoring stacks impose overhead incompatible with edge resource budgets. Yi proposed hierarchical monitoring for fog computing environments, distributing lightweight health checks across edge nodes with aggregated reporting to cloud-tier management platforms. Toczé and Nadjm-Tehrani examined the reliability implications of edge placement decisions, demonstrating that workload distribution across edge sites significantly affects both failure probability and diagnostic complexity.

More recently, adaptive monitoring strategies have been proposed to manage the tension between observability fidelity and resource consumption at the edge. Ren introduced dynamic monitoring granularity adjustment based on detected anomaly conditions, increasing collection frequency during potential incidents and reducing it during stable periods. However, their approach addressed single-site monitoring without cross-site or cross-tier correlation capabilities. The gap between edge-aware monitoring and full-spectrum telemetry correlation for fault diagnosis across edge-cloud boundaries remains largely unaddressed in literature.

D. AIOps and Autonomous Diagnostics

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

The AIOps paradigm, as surveyed by Dang and Notaro, applies machine learning across the incident lifecycle from anomaly detection through root-cause analysis to automated remediation. Within this paradigm, multi-modal learning approaches that jointly process heterogeneous telemetry signals have shown particular promise. Ma demonstrated that attention mechanisms capable of dynamically weighting different signal modalities outperform fixed-weight fusion for root-cause ranking. The integration of large language models for interpreting unstructured incident context has been explored by Ahmed, though their work focused on post-incident analysis rather than real-time diagnosis.

A critical challenge for AIOps in edge-cloud environments is the scarcity of labeled training data for edge-specific failure modes. Production edge deployments are relatively young compared to cloud infrastructure, and the diversity of edge hardware, network conditions, and application patterns generates a long tail of failure types that are poorly represented in existing incident databases. Semi-supervised and transfer learning strategies offer potential mitigations, but their application to edge-cloud fault diagnosis has received limited investigation.

III. METHODOLOGY / PROPOSED MODEL

A. Conceptual Framework

The CATCE architecture is grounded in a conceptual framework that decomposes telemetry correlation into four distinct analytical phases: signal normalization, contextual enrichment, adaptive fusion, and causal reasoning. These phases operate within a resource-proportional processing model that distributes analytical workload across the edge-cloud topology according to the computational and bandwidth resources available at each tier. The conceptual framework recognizes that effective correlation in edge-cloud systems requires not merely algorithmic sophistication but architectural accommodation of the fundamental asymmetries in connectivity, computing capacity, and telemetry fidelity that characterize these environments.

The framework models the edge-cloud system as a layered topology graph in which nodes represent computational entities (edge pods, cloud services, infrastructure components) and edges represent both logical dependencies (service call relationships) and physical connectivity (network links). Telemetry signals are treated as time-indexed observations attached to nodes and edges in this graph, and correlation is formulated as the problem of identifying subgraph structures whose associated telemetry patterns collectively indicate a localized fault condition. This graph-centric formulation naturally accommodates the topological reasoning required for fault isolation in distributed systems.

B. Semantic Enrichment Layer

The semantic enrichment layer addresses the heterogeneity challenge by transforming raw telemetry from diverse edge and cloud sources into a unified analytical representation. Each

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

telemetry signal is enriched with structural metadata derived from three sources: the deployment manifest (Kubernetes labels, namespace hierarchy, resource limits), the service mesh configuration (upstream and downstream dependencies, traffic policies, mutual TLS relationships), and the infrastructure topology (node placement, availability zone, network segment, edge site identity).

Enrichment occurs locally at each edge site and cloud region, avoiding the need to transmit raw telemetry to a central point for contextualization. The enriched telemetry carries a standardized attribute schema conforming to OpenTelemetry semantic conventions extended with edge-specific attributes (edge site identifier, WAN link identifier, last-mile network type). This schema standardization enables downstream correlation algorithms to reason across heterogeneous sources without per-source normalization logic, substantially simplifying the fusion and causal reasoning stages.

A notable design element is the inclusion of causal context tags that preserve the relationship between correlated events as they propagate through the system. When an edge agent detects a local anomaly, it attaches a context identifier to all subsequent telemetry emissions related to that anomaly. If the anomaly triggers cascading effects for example, an edge service experiencing latency due to a cloud backend degradation generating elevated error rates that propagate to dependent edge services the context identifier chains these events together, enabling the central correlation engine to reconstruct the causal sequence even when individual signals arrive out of order or with temporal gaps.

C. Adaptive Signal Fusion

The adaptive signal fusion mechanism determines how deeply and broadly to correlate telemetry signals based on two dynamic factors: the available computational budget at each processing tier and the estimated diagnostic significance of current conditions. During stable operation, fusion operates in an economic mode that computes lightweight statistical summaries and monitors pre-defined correlation indicators. When anomaly pre-screening detects potential issues, fusion transitions to an intensive mode that activates full cross-signal analysis, including temporal alignment of metric series with trace span timings, semantic matching of log event patterns with metric deviations, and network flow analysis correlated with application-level error bursts.

The computational budget is expressed as a resource envelope specifying the maximum CPU time, memory allocation, and network bandwidth that the fusion process may consume at each tier. Edge sites receive compact resource envelopes that constrain fusion to lightweight statistical methods; regional aggregation points receive moderate envelopes enabling cross-site pattern detection; the central cloud tier receives unrestricted envelopes supporting deep learning-based multi-modal analysis. This resource-proportional design ensures that correlation activities never compete with application workloads for scarce edge resources while still leveraging the full computational capacity of the cloud tier for complex diagnostic reasoning.

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

The fusion mechanism produces correlation bundles structured packages that encapsulate the set of correlated signals, their temporal and causal relationships, and a preliminary anomaly characterization derived from the fusion analysis. Correlation bundles serve as the input to the causal reasoning module and provide a natural unit of diagnostic context that can be transmitted between tiers, cached for postmortem analysis, or presented to human operators through incident management interfaces.

D. Topology-Guided Causal Reasoning

The causal reasoning module operates on correlation bundles augmented with the layered topology graph to produce fault isolation predictions. The reasoning process follows a constrained search strategy that leverages topological structure to prune the candidate root-cause space. Rather than considering all system components as potential root causes, the module restricts its search to components that are topologically reachable from the observed anomaly symptoms through plausible causal pathways in the dependency graph.

The causal scoring function integrates three evidence dimensions. Temporal evidence evaluates the consistency of anomaly onset timing with the expected propagation delays along each candidate causal pathway. Topological evidence assesses the structural plausibility of each candidate, favoring root causes connected to multiple observed symptoms through short dependency paths. Signal evidence measures the severity and distinctiveness of the telemetry anomaly at each candidate component, weighting candidates with more pronounced and characteristic anomaly signatures higher than those with marginal deviations.

A distinguishing feature of the reasoning module is its explicit handling of cross-tier causal chains. When symptoms are observed at edge sites, but the highest-scoring causal candidates reside in the cloud tier (or vice versa), the module applies a cross-tier propagation model that accounts for the WAN latency, bandwidth constraints, and protocol characteristics of the edge-to-cloud interconnection. This model adjusts temporal evidence calculations to reflect the expected delays introduced by cross-tier communication, preventing false temporal inconsistencies from disqualifying legitimate cross-tier root causes.

E. Resilience Under Network Degradation

The framework incorporates explicit resilience mechanisms for operating under the degraded network conditions that frequently characterize edge environments. A store-and-forward telemetry buffer at each edge site retains enriched telemetry during connectivity interruptions, transmitting accumulated data when connectivity is restored with timestamps and sequence numbers that enable the central engine to reconstruct the temporal ordering of events that occurred during the partition.

During network partitions, edge agents continue local anomaly detection and produce local diagnostic assessments with explicitly reduced confidence scores reflecting the absence of cross-

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

tier context. When connectivity resumes, a reconciliation protocol merges the edge-local assessments with central-tier analysis, updating fault isolation predictions based on the newly available cross-tier evidence. This resilience architecture ensures that edge-local reliability engineering continues to function during the network conditions that often accompany or cause the very failures that the system is designed to diagnose.

IV. EXPERIMENTAL SETUP

The evaluation testbed comprises a three-region cloud backbone (North America, Europe, Asia-Pacific) deployed on managed Kubernetes clusters totaling 32 worker nodes, and 18 edge sites distributed across metropolitan, suburban, and rural connectivity profiles implemented using lightweight K3s clusters with 2-3 worker nodes per site. The application workload models a distributed video analytics platform with 84 microservices edge services handle video ingestion, local inference, and content caching, while cloud services provide model management, global aggregation, user authentication, and persistent storage.

Edge-to-cloud connectivity profiles were configured to represent realistic deployment conditions: metropolitan sites operate with 50-100 Mbps links and 10-20 ms latency, suburban sites with 20-50 Mbps and 30-60 ms latency, and rural sites with 5-20 Mbps and 60-120 ms latency. The observability stack includes Prometheus, Jaeger, Fluentd, and a custom network flow exporter. CATCE components are deployed as sidecar operators at each tier.

A systematic fault injection campaign introduced 220 incidents across eight categories: edge pod failures (30), cloud service degradations (30), WAN impairments (30), edge resource saturation (25), cross-tier dependency timeouts (25), deployment regressions (25), certificate and authentication failures (20), and cascading multi-tier faults (35). Faults were parameterized with randomized severity, duration, and onset timing, and each scenario was repeated three times with different seeds for statistical robustness.

Three comparison systems were evaluated. Baseline C (centralized) collects all raw telemetry at the cloud tier for unified analysis. Baseline D (decentralized) runs independent anomaly detection at each edge site and cloud region with no cross-tier correlation. CATCE implements the full framework described in Section III. All three systems use identical anomaly detection algorithms to isolate the contribution of the correlation architecture from the choice of detection methodology. Fault isolation time was measured as the interval from fault injection to correct root-cause identification, and diagnostic precision was computed as the fraction of correctly identified root causes among all isolation attempts.

V. RESULTS AND ANALYSIS

A. Overall Fault Isolation Performance

CATCE achieved a mean fault isolation time of 3.8 minutes across all 220 injected incidents, compared to 13.2 minutes for Baseline C and 21.6 minutes for Baseline D reductions of 71.2 percent and 82.4 percent, respectively. The improvement was statistically significant at the $p < 0.001$ level using a paired Wilcoxon signed-rank test across all incident repetitions. Diagnostic precision was 94.6 percent for CATCE, 90.1 percent for Baseline C, and 72.3 percent for Baseline D.

The performance breakdown by failure tier revealed the expected interaction pattern. For edge-local failures such as pod crashes and resource saturation, CATCE and Baseline D achieved comparable detection latency (both under 2.5 minutes), as these failures produce strong local signals amenable to edge-only analysis. However, CATCE's diagnostic precision for edge-local failures was 97.1 percent compared to Baseline D's 88.4 percent, because CATCE's cross-tier context enabled it to rule out cloud-tier root causes and confirm edge-local diagnosis with higher confidence.

For cross-tier and cloud-originated faults dependency timeouts, deployment regressions, and cascading failures CATCE demonstrated its most substantial advantages. Mean isolation time for cascading multi-tier faults was 7.4 minutes with CATCE versus 19.8 minutes for Baseline C and 34.2 minutes for Baseline D. The causal context tags propagated through the semantic enrichment layer proved especially valuable for cascading faults, enabling the central reasoning module to reconstruct the cross-tier causal chain from correlated context identifiers rather than having to infer causal relationships purely from temporal and statistical patterns.

B. Performance Under Bandwidth Degradation

Bandwidth stress testing reduced all edge-to-cloud links to 20 percent of their baseline capacity. Under these conditions, Baseline C's mean isolation time increased by 163 percent to 34.7 minutes, as telemetry collection delays and aggressive sampling degraded diagnostic input quality. CATCE's mean isolation time increased by only 21 percent to 4.6 minutes, and diagnostic precision declined marginally from 94.6 to 92.8 percent. The adaptive fusion mechanism responded to bandwidth pressure by reducing the volume of raw telemetry escalated from edge to cloud tier while preserving the enriched summaries and causal context tags that carry the highest diagnostic information density per byte.

The store-and-forward buffer was exercised during simulated network partitions affecting 6 of the 18 edge sites. During partition periods averaging 8.4 minutes, edge agents continued producing local diagnostic assessments with appropriately reduced confidence. Upon reconnection, the reconciliation protocol successfully merged edge-local and central analyses within a mean of 47 seconds, upgrading or correcting local assessments based on cross-tier evidence. In 91 percent of partition-affected incidents, the final post-reconciliation diagnosis matched the ground truth, compared to 78 percent for edge-local assessments produced during the partition.

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

C. Signal Modality Contribution Analysis

Ablation experiments quantified the contribution of each telemetry signal type to overall diagnostic accuracy. Using all four modalities (metrics, traces, logs, network flow), CATCE achieved an F1 score of 0.946. Removing network flow data reduced F1 to 0.921, primarily affecting WAN impairment and network partition diagnosis. Removing logs reduced F1 to 0.908, with the largest impact on deployment regression and configuration error detection where error log patterns provide the most discriminating signal. Removing traces reduced F1 to 0.884, degrading cross-service latency attribution. Removing metrics produced the sharpest decline to 0.861, confirming that continuous metric series remain the most broadly informative single signal type for system health assessment.

The adaptive fusion mechanism's dynamic modality weighting was validated by comparing against fixed equal-weight fusion. The adaptive approach achieved a 4.7 percentage point F1 improvement over equal weighting, with the gain concentrated in failure categories where one modality carries disproportionate diagnostic value such as network flow data for WAN impairments and log patterns for configuration errors. This result confirms the theoretical motivation for learned modality weighting and demonstrates that the adaptive mechanism successfully identifies which signal types are most informative for each class of fault condition.

An additional finding of practical significance concerns the interaction between telemetry sampling and diagnostic accuracy. Under normal bandwidth conditions, CATCE's adaptive escalation transmitted approximately 8 percent of raw telemetry volume while achieving 94.6 percent diagnostic precision. Incrementally increasing the raw telemetry fraction to 15 and 25 percent improved precision to only 95.1 and 95.4 percent respectively, confirming that the semantic enrichment and causal context propagation mechanisms capture the vast majority of diagnostic information content within a compact representation. These diminishing returns curve provides practical guidance for bandwidth provisioning: reliability engineering teams can confidently allocate modest telemetry bandwidth budgets knowing that the enriched summary approach delivers near-maximal diagnostic value.

D. Resource Overhead Assessment

CATCE's edge agent consumed a mean of 0.38 CPU cores and 287 MB of memory per edge site, representing 6.2 percent of total edge compute capacity. This overhead remained stable across normal and anomaly-intensive periods, as the adaptive fusion mechanism manages computational load by adjusting analysis depth rather than spawning additional processes. Cloud-tier analysis consumed 4.1 percent of cloud cluster resources. Telemetry bandwidth consumption between edge and cloud tiers averaged 1.8 Mbps per site under normal conditions and peaked at 8.3 Mbps during intensive anomaly escalation, well within the capacity of even the most constrained rural edge links.

VI. DISCUSSION

A. Architectural Implications for Reliability Engineering

The experimental results demonstrate that the architectural choice of how and where to perform telemetry correlation has a far greater impact on diagnostic effectiveness than the specific algorithms employed for anomaly detection or root-cause ranking. All three evaluated systems shared identical detection algorithms, yet CATCE outperformed both baselines by wide margins. This finding suggests that reliability engineering teams seeking to improve their diagnostic capabilities in edge-cloud environments should prioritize investments in correlation architecture particularly the mechanisms for semantic enrichment, context propagation, and adaptive cross-tier information sharing over incremental improvements to detection model accuracy.

From an SRE practice perspective, the framework enables organizations to extend SLO-based reliability management to edge workloads with confidence that diagnostic capabilities will support the error budget governance and incident response workflows that SRE depends upon. The ability to maintain 92.8 percent diagnostic precision even under severely degraded bandwidth conditions means that edge sites need not be treated as observability blind spots requiring separate, less rigorous reliability management processes.

B. Integration with the Broader DevOps Ecosystem

CATCE's correlation bundles integrate naturally with adjacent DevOps practices. For shift-left testing, the bundle format provides a structured representation of failure propagation patterns that can be replayed in pre-production environments to validate observability coverage before deployment to edge sites. For DevSecOps, the integration with service mesh mutual TLS metadata enables correlation of security-relevant events certificate expirations, authentication failures, policy violations with their operational impact on service availability, bridging the gap between security monitoring and reliability engineering that often exists in zero-trust edge-cloud architectures.

FinOps considerations are addressed through the bandwidth-efficient design. The framework's telemetry compression and adaptive escalation reduce edge-to-cloud data transfer by approximately 89 percent compared to centralized full-fidelity collection. At production scale, this translates to substantial egress cost savings a material concern for organizations operating dozens or hundreds of edge sites whose telemetry transit costs can represent a significant fraction of total observability expenditure. The resource-proportional processing model further supports FinOps by ensuring that expensive cloud-tier analytical resources are engaged only when the diagnostic situation warrants, avoiding the constant computational overhead of running deep analysis on routine healthy-state telemetry.

C. Limitations and Validity Considerations

Several limitations should be acknowledged. The testbed operates on managed cloud and virtualized edge infrastructure that provides more predictable performance characteristics than

International Journal of Science, Technology and Convergence (IJSTC)
ISSN: 2134-986X

bare-metal edge deployments in challenging physical environments such as factory floors, remote energy installations, or mobile edge nodes in vehicular networks. The fault injection methodology, while comprehensive across eight categories, cannot capture the full diversity of production failure modes, particularly emergent failures arising from complex interactions among independently deployed microservice versions.

The causal reasoning module's effectiveness depends on the accuracy of the topology graph, which is constructed from deployment manifests and service mesh configuration. In environments where service interactions are not fully captured by the mesh such as systems using direct database connections, external API calls, or shared message queues not mediated by the service mesh the topology graph may be incomplete, degrading the quality of topological evidence used in causal scoring. Mechanisms for dynamic topology discovery through trace analysis could mitigate this limitation and represent an important direction for future development.

Clock synchronization accuracy, while generally sufficient for the temporal correlation granularity required by the causal reasoning module, exhibited degradation during network congestion periods at rural edge sites where NTP accuracy deteriorated to tens of milliseconds. For rapidly propagating failures where the causal chain traverses multiple hops within milliseconds, this synchronization uncertainty introduces noise into temporal evidence calculations. The deployment of Precision Time Protocol where supported, or the adoption of vector clock-based logical ordering as a complement to physical timestamps, would address this limitation.

VII. CONCLUSION AND FUTURE WORK

This paper has presented the Context-Aware Telemetry Correlation Engine, a framework that addresses the fundamental observability challenges of edge-cloud systems through semantic enrichment, adaptive signal fusion, and topology-guided causal reasoning. By distributing correlation intelligence proportionally across the edge-cloud topology and preserving causal context through a dedicated propagation protocol, CATCE achieves real-time fault isolation performance that substantially exceeds both centralized and decentralized baselines while operating within the resource and bandwidth constraints characteristic of production edge deployments.

The experimental evaluation demonstrated a 71.2 percent reduction in mean fault isolation time, diagnostic precision of 94.6 percent under normal conditions and 92.8 percent under severe bandwidth degradation, and resource overhead compatible with production edge constraints. The ablation analysis confirmed the value of multi-modal signal fusion with adaptive modality weighting, while the bandwidth stress tests validated the framework's resilience design under the degraded conditions that frequently accompany edge-cloud failures.

Future work will pursue several complementary directions. First, the incorporation of foundation models for interpreting unstructured incident context deployment changelogs, configuration diffs, and communication artifacts could enhance causal reasoning for novel failure types underrepresented in historical training data. Second, the application of continual learning techniques to the adaptive fusion mechanism would enable the system to refine its modality weighting and escalation thresholds based on evolving operational patterns without requiring periodic retraining. Third, privacy-preserving correlation techniques based on secure multi-party computation or differential privacy would enable cross-organizational telemetry correlation in multi-tenant edge infrastructure where tenants require confidentiality guarantees.

Fourth, the extension of the resource-proportional processing model to heterogeneous accelerator hardware including edge AI inference chips and FPGA-based network processors could unlock additional analytical capabilities at the edge tier without increasing general-purpose compute consumption. Finally, longitudinal deployment studies in production edge-cloud environments spanning diverse industry verticals would provide essential validation of the framework's robustness, maintainability, and operational value under the non-stationarity and continuous evolution that characterize real-world distributed systems.

REFERENCES

[1] B. H. Sigelman, L. A. Barroso, M. Burrows, P. Stephenson, M. Plakal, D. Beaver, S. Jaspan, and C. Shanbhag, "Dapper, a large-scale distributed systems tracing infrastructure," Google, Inc., Tech. Rep., 2010.

[2] OpenTelemetry Authors, "OpenTelemetry specification," Cloud Native Computing Foundation, 2023. [Online]. Available: https://opentelemetry.io/docs/specs/otel/

[3] R. R. Sambasivan, R. Fonseca, I. Shafer, and G. R. Ganger, "So, you know that your latency is not normal," ACM Queue, vol. 11, no. 12, pp. 50-63, 2013.

[4] J. Kaldor, J. Mace, M. Bejda, E. Gao, W. Kuropatwa, J. O'Neill, K. W. Ong, B. Schreiber, P. Sajber, S. Balakrishnan, and K. Karr, "Canopy: An end-to-end performance tracing and analysis system," in Proc. ACM Symp. Operating Systems Principles, Shanghai, China, 2017, pp. 34-50.

[5] B. Beyer, C. Jones, J. Petoff, and N. R. Murphy, Site Reliability Engineering: How Google Runs Production Systems. Sebastopol, CA, USA: O'Reilly Media, 2016.

[6] W. Xu, L. Huang, A. Fox, D. Patterson, and M. I. Jordan, "Detecting large-scale system problems by mining console logs," in Proc. ACM SIGOPS 22nd Symp. Operating Systems Principles, Big Sky, MT, USA, 2009, pp. 117-132.

[7] P. Aggarwal, J. Sahoo, J. Moss, and R. Kompella, "On unified telemetry data analysis for cloud infrastructure fault localization," IEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 4023-4036, 2022.

[8] P. Chen, Y. Qi, P. Zheng, and D. Hou, "CauseInfer: Automatic and distributed performance diagnosis with hierarchical causality graph in large distributed systems," in Proc. IEEE INFOCOM, Toronto, ON, Canada, 2014, pp. 1887-1895.

[9] L. Wu, J. Tordsson, E. Elmroth, and O. Kao, "MicroRCA: Root cause localization of performance issues in microservices," in Proc. IEEE/IFIP Network Operations and Management Symp., Budapest, Hungary, 2020, pp. 1-9.

[10] Y. Meng, S. Zhang, Y. Sun, R. Zhang, Z. Hu, Y. Zhang, C. Jia, Z. Wang, and D. Pei, "Localizing failure root causes in a microservice through causality inference," in Proc. IEEE/ACM 28th Int. Symp. Quality of Service, Hang Zhou, China, 2020, pp. 1-10.

[11] Z. Li, Y. Chen, D. Sui, S. Yang, and T. Zhang, "Transfer entropy-based fault propagation analysis for microservice systems," IEEE Transactions on Services Computing, vol. 16, no. 3, pp. 1987-1999, 2023.

[12] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, 2016.

[13] S. Yi, Z. Hao, Z. Qin, and Q. Li, "Fog computing: Platform and applications," in Proc. IEEE Workshop on Hot Topics in Web Systems and Technologies, Washington, DC, USA, 2015, pp. 73-78.

[14] K. Toczé and S. Nadjm-Tehrani, "A taxonomy for management and optimization of multiple resources in edge computing," Wireless Communications and Mobile Computing, vol. 2018, pp. 1-23, 2018.

[15] X. Ren, P. London, J. Ziani, and A. Wierman, "Adaptive monitoring for edge-cloud systems," in Proc. IEEE Int. Conf. Cloud Computing, San Jose, CA, USA, 2021, pp. 112-121.

[16] Y. Dang, Q. Lin, and P. Huang, "AIOps: Real-world challenges and research innovations," in Proc. IEEE/ACM 41st Int. Conf. Software Engineering: Companion, Montreal, QC, Canada, 2019, pp. 4-7.

[17] P. Notaro, J. Cardoso, and M. Gerndt, "A systematic mapping study in AIOps," in Proc. Int. Conf. Service-Oriented Computing, Dubai, UAE, 2021, pp. 110-123.

[18] M. Ma, W. Lin, D. Pan, and P. Wang, "MS-Rank: Multi-metric and multi-source based root cause ranking for cloud service failure diagnosis," IEEE Transactions on Services Computing, vol. 15, no. 6, pp. 3550–3563, 2022.

[19] T. Ahmed, H. Yin, and R. White, "Recommending root-cause and mitigation steps for cloud incidents using large language models," in Proc. IEEE/ACM 45th Int. Conf. Software Engineering, Melbourne, Australia, 2023, pp. 1314-1326